(19)日本国特許庁（ＪＰ）　(12)**公開特許公報（Ａ）**　(11)特許出願公開番号

特開平5－346797

(43)公開日　平成５年(1993)12月27日

(51)Int.Cl.[5]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　9/00		Ｃ　8946－5Ｈ		
Ｇ０６Ｆ　15/36		Ｄ　8320－5Ｌ		
Ｇ１０Ｌ　3/00	５１３	Ｄ　8842－5Ｈ		
	５１５	Ｄ　8842－5Ｈ		

審査請求　未請求　請求項の数８（全 20 頁）

(21)出願番号　特願平5－828

(22)出願日　平成５年(1993)１月６日

(31)優先権主張番号　特願平4－121460

(32)優先日　平４(1992)４月15日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000002185
ソニー株式会社
東京都品川区北品川６丁目７番35号

(72)発明者　西口　正之
東京都品川区北品川６丁目７番35号　ソニー株式会社内

(72)発明者　松本　淳
東京都品川区北品川６丁目７番35号　ソニー株式会社内

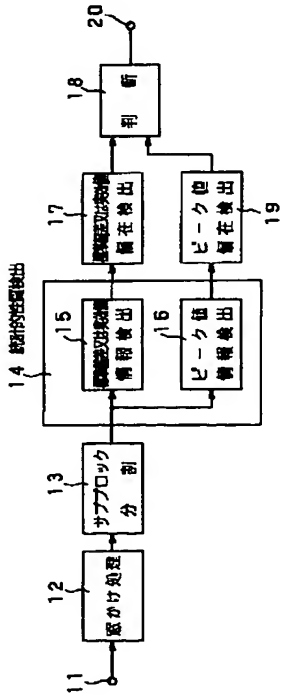
(74)代理人　弁理士　小池　晃　（外２名）

(54)【発明の名称】　有声音判別方法

(57)【要約】

【構成】　入力信号の１ブロック（１フレーム）をサブブロック分割部１３でさらに分割し、統計的性質検出部１４でサブブロック毎に標準偏差又は実効値の情報とピーク値情報とを検出する。偏在検出部１７では標準偏差又は実効値の時間軸上での偏在を検出し、偏在検出部１９ではピーク値の偏在を検出して、これらの偏在検出情報に基づいて判断部１８が各ブロック毎に有声音か無声音かを判別する。

【効果】　有声音か無声音（又はノイズ）かの判別が確実に行える。



1

【特許請求の範囲】

【請求項１】　入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、

１ブロックの信号を複数のサブブロックに分割する工程と、

上記複数のサブブロック毎に信号の統計的な性質を求める工程と、

上記統計的な性質の時間軸上での偏りに応じて有声音か否かを判別する工程とを有することを特徴とする有声音判別方法。

【請求項２】　上記信号の統計的な性質とは各サブブロック毎の信号のピーク値、実効値又は標準偏差であることを特徴とする請求項１記載の有声音判別方法。

【請求項３】　入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、

１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、

上記１ブロックの信号のレベルを求める工程と、

上記１ブロックの信号の周波数軸上のエネルギー分布と信号レベルとに応じて有声音か否かを判別する工程とを有することを特徴とする有声音判別方法。

【請求項４】　入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、

１ブロックの信号を複数のサブブロックに分割する工程と、

上記複数のサブブロック毎の信号のピーク値、実効値又は標準偏差の時間軸上での偏りを求める工程と、

１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、

上記１ブロックの信号のレベルを求める工程と、

上記複数のサブブロック毎の信号のピーク値、実効値又は標準偏差の時間軸上での偏りと上記１ブロックの信号の周波数軸上のエネルギー分布又は上記１ブロックの信号のレベルとに応じて有声音か否かを判別する工程とを有することを特徴とする有声音判別方法。

【請求項５】　入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、

１ブロックの信号を複数のサブブロックに分割する工程と、

上記複数のサブブロック毎に時間軸上で信号のピーク値、実効値又は標準偏差を求める工程と、

上記１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、

上記１ブロックの信号のレベルを求める工程と、

上記複数のサブブロック毎の信号のピーク値、実効値又は標準偏差と上記１ブロックの信号の周波数軸上のエネ

2

ルギー分布と上記１ブロックの信号のレベルとに応じて有声音か否かを判別する工程とを有することを特徴とする有声音判別方法。

【請求項６】　入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、

１ブロックの信号を複数のサブブロックに分割する工程と、

上記複数のサブブロック毎に時間軸上で信号の実効値を求め、この実効値の標準偏差と平均値とに基づいてサブブロック毎の実効値の分布を求める工程と、

上記１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、

上記１ブロックの信号のレベルを求める工程と、

上記複数のサブブロック毎の実効値の分布と上記１ブロックの信号の周波数軸上のエネルギー分布と上記１ブロックの信号のレベルとの少なくとも２つに応じて有声音か否かを判別する工程とを有することを特徴とする有声音判別方法。

【請求項７】　上記複数のサブブロック毎の実効値の分布と上記１ブロックの信号の周波数軸上のエネルギー分布と上記１ブロックの信号のレベルとの少なくとも１つの時間的な変化をトラッキングし、その結果に基づいて有声音か否かを判別することを特徴とする請求項６記載の有声音判別方法。

【請求項８】　上記１ブロックの信号について複数の周波数バンド毎に有声音／無声音の識別フラグを設定する際に、上記有声音判別工程において否と判別されたブロックは、全てのバンドを無声音フラグとすることを特徴とする請求項６記載の有声音判別方法。

【発明の詳細な説明】

【０００１】

【産業上の利用分野】本発明は、音声信号から有声音を雑音又は無声音と区別して判別する有声音判別方法に関する。

【０００２】

【従来の技術】音声は音の性質として有声音と無声音に区別される。有声音は声帯振動を伴う音声で周期的な振動として観測される。無声音は声帯振動を伴わない音声で非周期的な音として観測される。通常の音声では大部分が有声音であり、無声音は無声子音と呼ばれる特殊な子音のみである。有声音の周期は声帯振動の周期で決まり、これをピッチ周期、その逆数をピッチ周波数という。これらピッチ周期及びピッチ周波数（以下、ピッチとした場合はピッチ周期を指す）は声の高低やイントネーションを決める重要な要因である。したがって、上記ピッチをどれだけ正確に捉えるかが音声の音質を左右する。しかし、上記ピッチを捉える場合には、上記音声の周囲にある雑音いわゆる背景雑音や量子化の際の量子化雑音を考慮しなければならない。これらの雑音又は無声

音と有声音を区別することが音声信号を符号化する場合に重要となる。

【０００３】上記音声信号の符号化の具体的な例としては、ＭＢＥ（Multiband Excitation: マルチバンド励起）符号化、ＳＢＥ（Singleband Excitation:シングルバンド励起）符号化、ハーモニック（Harmonic）符号化、ＳＢＣ（Sub-band Coding:帯域分割符号化）、ＬＰＣ（Linear Predictive Coding: 線形予測符号化）、あるいはＤＣＴ（離散コサイン変換）、ＭＤＣＴ（モデファイドＤＣＴ）、ＦＦＴ（高速フーリエ変換）等がある。

【０００４】例えば、上記ＭＢＥ符号化においては、入力音声信号波形からピッチを抽出する場合、明確なピッチが表れない場合でもピッチの軌跡を捉えやすくしていた。そして、復号化側（合成側）は、上記ピッチを基に余弦波（cosin)波合成により時間軸上の有声音波形を合成し、別途合成される時間軸上の無声音波形と加算合成し出力する。

【０００５】

【発明が解決しようとする課題】ところで、ピッチを捉えやすくすると上記背景雑音等の部分で本来のピッチでない間違ったピッチを捉えてしまう場合がある。もし、上記ＭＢＥ符号化で間違ったピッチを捉えてしまうと、合成側では、その間違ったピッチの所で各cosin波のピークが重なるようにcosin 波合成を行ってしまう。すなわち、誤って捉えたピッチ周期毎に有声音の合成で行っているような固定位相（０位相又はπ／２位相）の加算で各cosin 波を合成し、ピッチが得られない筈の背景雑音等を周期性を持つインパルス波形として合成する。つまり、本来、時間軸上で散らばっているべき背景雑音等の振幅の強度があるフレームの１部分に周期性を持ちながら集中してしまい、非常に耳障りな異音を再生してしまうことになる。

【０００６】本発明は、上記実情に鑑みてなされたものであり、有声音を雑音又は無声音と区別し確実に判別でき、合成側に対しては異音の発生を抑えさせることができる有声音判別方法の提供を目的とする。

【０００７】

【課題を解決するための手段】本発明に係る有声音判別方法は、入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、１ブロックの信号を複数のサブブロックに分割する工程と、上記複数のサブブロック毎に信号の統計的な性質を求める工程と、上記統計的な性質の時間軸上での偏りに応じて有声音か否かを判別する工程とを有することを特徴として上記課題を解決することができる。

【０００８】ここで、上記信号の統計的な性質には、各サブブロック毎の信号のピーク値、実効値又は標準偏差を用いることができる。

【０００９】他の発明に係る有声音判別方法として、入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、上記１ブロックの信号のレベルを求める工程と、上記１ブロックの信号の周波数軸上のエネルギー分布と信号レベルとに応じて有声音か否かを判別する工程とを有することを特徴として上記課題を解決することができる。

【００１０】ここで、上記各サブブロック毎の信号のピーク値、実効値又は標準偏差という統計的な性質と上記１ブロックの信号の周波数軸上のエネルギー分布とに応じて又は上記各サブブロック毎の信号のピーク値、実効値又は標準偏差という統計的な性質と上記１ブロックの信号のレベルとに応じて有声音か否かを判別してもよい。

【００１１】さらに他の発明に係る有声音判別方法として、入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、１ブロックの信号を複数のサブブロックに分割する工程と、上記複数のサブブロック毎に時間軸上で信号のピーク値、実効値又は標準偏差を求める工程と、上記１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、上記１ブロックの信号のレベルを求める工程と、上記複数のサブブロック毎の信号のピーク値、実効値又は標準偏差と上記１ブロックの信号の周波数軸上のエネルギー分布と上記１ブロックの信号のレベルとに応じて有声音か否かを判別する工程とを有することを特徴として上記課題を解決することができる。

【００１２】またさらに他の発明に係る有声音判別方法として、入力された音声信号をブロック単位で分割して各ブロック毎に有声音か否かの判別を行う有声音判別方法において、１ブロックの信号を複数のサブブロックに分割する工程と、上記複数のサブブロック毎に時間軸上で信号の実効値を求め、この実効値の標準偏差と平均値とに基づいてサブブロック毎の実効値の分布を求める工程と、上記１ブロックの信号の周波数軸上のエネルギー分布を求める工程と、上記１ブロックの信号のレベルを求める工程と、上記複数のサブブロック毎の実効値の分布と上記１ブロックの信号の周波数軸上のエネルギー分布と上記１ブロックの信号のレベルとの少なくとも２つに応じて有声音か否かを判別する工程とを有することを特徴としている。

【００１３】ここでいう有声音か否かの判別とは、有声音か雑音又は無声音かを判別することであり、有声音を確実に判別すると共に雑音又は無声音も確実に判別できる。つまり、入力音声信号から雑音（背景雑音）又は無声音を判別することもできる。このようなときには、例えば、強制的に入力音声信号の全帯域を無声音とすると、合成側での異音の発生を抑えることができる。

【００１４】

【作用】有声音と雑音又は無声音の統計的な性質の時間軸上で偏りが異なるため、入力音声信号が有声音か雑音又は無声音であるかを判別することができる。

【００１５】

【実施例】以下、本発明に係る有声音判別方法の実施例について、図面を参照しながら説明する。図１は、本発明の第１の実施例となる有声音判別方法を説明するための有声音判別装置の概略構成を示している。この第１の実施例は、音声の１ブロックの信号をさらに分割したサブブロック毎の信号の統計的な性質の時間軸上での偏りに応じて有声音か否かを判別する。

【００１６】図１において、入力端子１１には、図示しないＨＰＦ（ハイパスフィルタ）等のフィルタによりいわゆるＤＣ（直流）オフセット分の除去や帯域制限（例えば２００～３４００Hzに制限）のための少なくとも低域成分（２００Hz以下）の除去が行われた音声の信号が供給される。この信号は、窓かけ処理部１２に送られる。この窓かけ処理部１２では１ブロックＮサンプル（例えばＮ＝２５６）に対して方形窓をかけ、この１ブロックを１フレームＬサンプル（例えばＬ＝１６０）の間隔で時間軸方向に順次移動させており、各ブロック間のオーバーラップはＮ－Ｌサンプル（９６サンプル）となっている。上記窓かけ処理部１２からのＮサンプルのブロックの信号は、サブブロック分割部１３に供給される。このサブブロック分割部１３は、上記窓かけ処理部１２で分割された１ブロックの信号をさらに細分割する。そして、得られたサブブロック毎の信号は、統計的性質検出部１４に供給される。この統計的性質検出部１４は、本第１の実施例の場合、標準偏差又は実効値情報検出部１５及びピーク値情報検出部１６からなる。上記標準偏差又は実効値情報検出部１５で得られた標準偏差又は実効値情報は、標準偏差又は実効値偏在検出部１７に供給される。この標準偏差又は実効値偏在検出部１７は、標準偏差又は実効値情報から時間軸上での偏りを検出する。そして、この時間軸上での標準偏差又は実効値の偏在情報は、判断部１８に供給される。この判断部１８は、時間軸上での標準偏差又は実効値の偏在情報を例えば所定の閾値と比較することよりサブブロック毎の信号が有声音であるか否かを判断し、その情報を出力端子２０から導出する。一方、上記ピーク値情報検出部１６で得られたピーク値情報は、ピーク値偏在検出部１９に供給される。このピーク値偏在検出部１９は、上記ピーク値情報から時間軸上での信号のピーク値の偏りを検出する。そして、この時間軸上での信号のピーク値の偏在情報は、判断部１８に供給される。この判断部１８は、上記時間軸上での信号のピーク値の偏在情報を例えば所定の閾値と比較することによりサブブロック毎の信号が有声音であるか否かを判断し、その判断情報を出力端子２０から導出する。

【００１７】次に、本第１の実施例で統計的性質として用いられる各サブブロック毎の信号のピーク値情報、標準偏差又は実効値情報の検出とそれらの時間軸上での偏在の検出について説明する。

【００１８】ここで、上記各サブブロック毎の信号のピーク値、標準偏差又は実効値を本第１の実施例で用いるのは、有声音と雑音又は無声音の信号のピーク値、標準偏差又は実効値が時間軸上で著しく異なるためである。例えば、図２のＡに示すような音声の母音（有声音）と図２のＣに示すような雑音又は子音（無声音）を比較する。母音の振幅のピークの並びは、図２のＡのように時間軸上で偏りながらも規則的であるのに対し、雑音又は子音の振幅のピークの並びは時間軸上で一様（フラット）であるが不規則である。また、母音の標準偏差又は実効値も、図２のＢに示すように時間軸上で偏っているのに対し、雑音又は子音の標準偏差又は実効値は、図２のＤに示すように時間軸上でフラットである。

【００１９】先ず、信号の上記各サブブロック毎の標準偏差又は実効値情報を検出する標準偏差又は実効値情報検出部１５と該標準偏差又は実効値情報の時間軸上での偏在の検出について説明する。この標準偏差又は実効値情報検出部１５は、図３に示すように入力端子２１からのサブブロック毎の信号から標準偏差又は実効値を算出する標準偏差又は実効値算出部２２と、該標準偏差又は実効値から相加平均を算出する相加平均算出部２３と、上記標準偏差又は実効値から相乗平均値を算出する相乗平均算出部２４とからなる。そして、上記相加平均値と相乗平均値より時間軸上での偏在情報を標準偏差又は実効値偏在検出部１７が検出し、判断部１８が該偏在情報からサブブロック毎の音声信号が有声音か否かを判断し、その判断情報が出力端子２０から導出される。

【００２０】上記エネルギーの分散から有声音か否かを判断する原理を図１と図３を用いて説明する。上記窓かけ処理部１２で方形窓をかけることにより切り出される１ブロックのサンプル数Ｎを２５６サンプルとし、入力サンプル列を$x(n)$とする。この１ブロック（２５６サンプル）を上記サブブロック分割部１３により８サンプル毎に分割する。するとサブブロック長B_l＝８のサブブロックがN/B_l（２５６／８＝３２）個上記１ブロックの中に存在することになる。この３２個のサブブロック毎の時間軸上データは、上記標準偏差又は実効値情報検出部１５の例えば標準偏差又は実効値算出部２２に供給される。

【００２１】この標準偏差又は実効値算出部２２は、上記３２個のサブブロック毎に時間軸上データの例えば標準偏差$\sigma_a(i)$として、

【００２２】

【数１】

7

$$\sigma_a(i)=\sqrt{\frac{1}{B_l}\sum_{n=k}^{k+B_l-1}(x(n)-\underline{x})^2} \qquad 0\leqq i<N/B_l$$

$$k=i\times B_l \qquad 0\leqq i<N/B_l$$

8

… （1）

【００２３】で示される（１）式により算出した値を出力する。ここでｉはサブブロックのインデックスであり、ｋはサンプル数である。また、$\underline{x}$は１ブロック当たりの入力サンプルの平均値である。この平均値$\underline{x}$は、１ブロックの全サンプル（Ｎ個）の平均であり、各サブブロック毎の平均ではないことに注意すべきである。

【００２４】また、上記サブブロック毎の実効値は、上記（１）式中の$(x(n)-\underline{x})^2$の代わりに、各サンプルｘについて上記１ブロック内のサンプルの平均値$\underline{x}$との差をとらない$(x(n))^2$を用いたものであり、いわゆるｒｍｓ（root meansquare、自乗平均の平方根）とも称されるものである。

【００２５】上記標準偏差$\sigma_a(i)$は、時間軸上での分散を調べるために上記相加平均算出部２３及び相乗平均算出部２４に供給される。上記相加平均算出部２３及び相乗平均算出部２４は、相加平均値$a_{v:add}$及び相乗平均値$a_{v:mpy}$を、

【００２６】

【数２】

$$a_{v:add}=\frac{1}{N/B_l}\sum_{i=0}^{N/B_l-1}\sigma_a(i) \qquad \cdots (2)$$

$$a_{v:mpy}=\left\{\prod_{i=0}^{N/B_l-1}\sigma_a(i)\right\}^{1/N/B_l} \qquad \cdots (3)$$

【００２７】で示される（２）及び（３）式により算出する。これらの（１）式～（３）式では標準偏差についてのみ例示しているが、実効値の場合も同様であることは勿論である。

【００２８】上記（２）及び（３）式により算出された

$$p_f=a_{v:add}/a_{v:mpy}$$

で求める。この比率p_fは、時間軸上の標準偏差の偏在を表す偏在情報である。この偏在情報（比率）p_fは、判断部１８に供給され、該判断部１８では、例えば、上記偏在情報p_fを閾値p_{thf}と比較し有声音か否かの判断を行う。例えば、上記閾値p_{thf}を1.1に設定しておき、上記偏在情報p_fが該閾値p_{thf}より大きいと標準偏差又は実効値の偏りが大きいと判断し有声音とする。一方、上記分散情報p_fが該閾値p_{thf}より小さいと標準偏差又は実効値の偏りが小さい（フラットである）と判断し有声音でない（雑音又は無声音である）とする。

【００２９】次に、ピーク値情報を検出するピーク値情報検出部１６と該ピーク値の時間軸上での偏在の検出について説明する。このピーク値情報検出部１６は、図４に示すように入力端子２１からのサブブロック毎の信号からピーク値を検出するピーク値検出部２６と、このピーク値検出部２６からのピーク値の平均値を算出する平均ピーク値算出部２７と、入力端子２５を介して供給されるブロック毎の信号から標準偏差値を算出する標準偏差算出部２８とからなる。そして、上記ピーク値偏在検出部１９が上記平均ピーク値算出部２７からの平均ピーク値を上記標準偏差算出部２８からのブロック毎の標準偏差値で除算し、時間軸上での平均ピーク値の偏在を検出する。この平均ピーク値偏在情報は、判断部１８に供給される。この判断部１８が該平均ピーク値偏在情報を基にサブブロック毎の音声信号が有声音か否かを判断し、該判断情報が出力端子２０から導出される。

【００３０】上記ピーク値情報から有声音か否かを判断する原理を図１と図４を用いて説明する。上記ピーク値検出部２６には、上記窓かけ処理部１２、サブブロック分割部１３及び入力端子２１を介してサブブロック長B_l（例えば８）のサブブロック分の信号がN/B_l（２５６／８＝３２）個供給される。このピーク値検出部２６は、例えば３２個分のサブブロック毎のピーク値Ｐ(i)を、

【００３１】

【数３】

相加平均値$a_{v:add}$及び相乗平均値$a_{v:mpy}$は、上記標準偏差又は実効値偏在検出部１７に供給される。この標準偏差又は実効値偏在検出部１７は、上記相加平均値$a_{v:add}$と相乗平均値$a_{v:mpy}$とから比率p_fを、

… （4）

9　　10

$$P(i) = \underset{k \leq n \leq k+B_1-1}{MAX} (|x(n)|) \qquad 0 \leq i < N/B_1$$

$$k = i \times B_1$$

… (5)

【0032】で示される（5）式の条件で検出する。ここでiはサブブロックのインデックスであり、kはサンプル数である。また、MAXは最大値を求める関数である。

$$\underline{P} = \frac{1}{N/B_1} \sum_{i=0}^{N/B_1-1} P(i)$$

【0035】で示される（6）式により算出する。

【0036】また、上記標準偏差算出部28は、ブロック毎の標準偏差値σ_b（i）を、

$$\sigma_b(i) = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x(n) - \underline{x})^2} \qquad 0 \leq i < N$$

… (7)

【0038】で求める。そして、上記ピーク値偏在検出部19は、ピーク値偏在情報P_nを上記平均ピーク値$\underline{P}$

$$P_n = \underline{P} / \sigma_b(i)$$

のように算出する。なお、上記標準偏差算出部28の代わりに、実効値（rms値）を算出する実効値算出部を用いてもよい。

【0039】上記（8）式により算出されたピーク値偏在情報P_nは、時間軸上でのピーク値の偏在の度合いを示すもので、上記判断部18に供給される。そして、上記判断部18は、例えば、上記ピーク値偏在情報P_nを閾値P_{thn}と比較し有声音か否かの判断を行う。例えば、上記、ピーク値偏在情報P_nが該閾値P_{thn}より大きいとピーク値の時間軸上での偏りが大きいと判断し有声音とする。一方、上記ピーク値偏在情報P_nが閾値P_{thn}より小さいとピーク値の偏りが小さいと判断し有声音でない（雑音又は無声音である）とする。

【0040】以上により、本発明に係る有声音判別方法の第1の実施例は、各サブブロック毎の信号のピーク値、実効値又は標準偏差のような統計的性質の時間軸上での偏りに応じて有声音か否かを判別することができる。

【0041】次に図5は、本発明の第2の実施例としての有声音判別方法を説明するための有声音判別装置の概略構成を示す図である。この第2の実施例は、音声の1ブロックの信号の周波数軸上のエネルギーの分布とレベルとから有声音か否かを判別する。

【0042】この第2の実施例は、有声音のエネルギー分布が周波数軸上の低域側に集中し、雑音又は無声音のエネルギー分布が周波数軸上の高域側に集中する傾向を用いている。

【0043】この図5において、入力端子31には、図示しないHPF（ハイパスフィルタ）等のフィルタによりいわゆるDC（直流）オフセット分の除去や帯域制限（例えば200～3400Hzに制限）のための少なくとも低域成分（200Hz以下）の除去が行われた音声の信号が供給される。この信号は、窓かけ処理部32に送られる。この窓かけ処理部32では1ブロックNサンプル（例えばN＝256）に対して例えばハミング窓をかけ、この1ブロックを1フレームLサンプル（例えばL＝160）の間隔で時間軸方向に順次移動させており、各ブロック間のオーバーラップはN－L（96サンプル）となっている。この窓かけ処理部32でNサンプルのブロックとされた信号は、直交変換部33に供給される。この直交変換部33は、例えば1ブロック256サンプルのサンプル列に対して1792サンプル分の0データを付加して（いわゆる0詰めして）2048サンプルとし、この2048サンプルの時間軸データ列に対して、FFT（高速フーリエ変換）等の直交変換処理を施し、周波数軸データ列に変換する。この直交変換部33からの周波数軸上のデータは、エネルギー検出部34に供給される。このエネルギー検出部34は、供給された周波数軸上データを低域側と高域側に分け、それぞれ低域側エネルギー検出部34aと高域側エネルギー検出部34bによりエネルギーを検出する。この低域側エネルギー検出部34a及び高域側エネルギー検出部34bにより検出された低域側エネルギー検出値及び高域側エネルギー検出値は、エネルギー分布算出部35に供給され、比率（エネルギー分布情報）が求められる。このエネルギー分布算出部35により求められたエネルギー分

【0033】そして、上記平均ピーク値算出部27が上記ピーク値P(i) から平均ピーク値$\underline{P}$を、

【0034】

【数4】

… (6)

【0037】

【数5】

… (8)

11

布情報は、判断部３７に供給される。また、上記低域側エネルギー検出値と高域側エネルギー検出値は、信号レベル算出部３６に供給され、１サンプル当たりの信号のレベルが計算される。この信号レベル算出部３６によって算出された信号レベル情報は、上記判断部３７に供給される。上記判断部３７は、上記エネルギー分布情報及び信号レベル情報を基に入力音声信号が有声音であるか否かを判断し、判断情報を出力端子３８から導出する。

【００４４】以下に、この第２の実施例の動作を説明する。上記窓かけ処理部３２でハミング窓をかけることに

$$a_m(j) = \sqrt{R_e^2(j) + I_m^2(j)} \quad \cdots (9)$$

【００４６】により求められる。この（９）式でＲｅ（ｊ）は実数部を表し、Ｉｍ（ｊ）は虚数部を表す。また、ｊはサンプル数で０以上Ｎ／２（＝１２８サンプル）未満の範囲にある。

【００４７】上記エネルギー検出部３４の低域側エネル

$$S_L = \sum_{j=0}^{N/4-1} a_m^2(j) \quad \cdots (10)$$

$$S_H = \sum_{j=N/4}^{N/2-1} a_m^2(j) \quad \cdots (11)$$

【００４９】で示される（10）式及び(11)式により求める。ここでいう低域側は０～２KHz、高域側は２～3.4KHzの周波数帯である。上記(10)、(11)式により算出された低域側エネルギーＳＬ及び高域側エネルギーＳＨは

$$f_b = S_L / S_H \quad \cdots (12)$$

となる。

【００５０】この周波数軸上でのエネルギー分布情報ｆｂは、判断部３７に供給される。この判断部３７は、上記エネルギー分布情報ｆｂを例えば閾値ｆthbと比較し有声音か否かの判断を行う。例えば上記閾値ｆthbを１５に設定しておき上記エネルギー分布情報ｆｂが該閾値ｆthbより小さいときは高域側にエネルギーが集中していて有声音でない（雑音又は無声音である）確率が高い

$$\ell_a = \sqrt{\frac{S_L + S_H}{N/2}} \quad \cdots (13)$$

【００５３】で示される（13）式から求める。この平均レベル情報ｌａも判断部３７に供給される。この判断部３７は、上記平均レベル情報ｌａを例えば閾値ｌthaと比較し有声音か否かの判断を行う。例えば上記閾値ｌthaを550に設定しておき上記平均レベル情報ｌａが該閾値ｌthaより小さいときは有声音でない（雑音又は無声音である）確率が高いと判断することになる。

【００５４】上記判断部３７は、上記エネルギー分布情

12

より切り出される１ブロックのサンプル数Ｎを２５６サンプルとし、入力サンプル列をx(n)とする。この１ブロック（２５６サンプル）の時間軸上のデータは、上記直交変換部３３により１ブロックの周波数軸上のデータに変換される。この１ブロックの周波数軸上のデータは、上記エネルギー検出部３４に供給され、振幅ａｍ（ｊ）が、

【００４５】

【数６】

ギー検出部３４ａ及び高域側エネルギー検出部３４ｂでは、上記（９）式に示された振幅ａｍ（ｊ）から、低域側エネルギーＳＬ及び高域側エネルギーＳＨ及びを、

【００４８】

【数７】

上記分布算出部３５に供給され、その比率ＳＬ／ＳＨにより周波数軸上でのエネルギーの分布のバランス情報（エネルギー分布情報）ｆｂが求められる。すなわち、

と判断することになる。

【００５１】また、上記低域側エネルギーＳＬ及び高域側エネルギーＳＨは、上記信号レベル算出部３６に供給される。この信号レベル算出部３６は、上記低域側エネルギーＳＬ及び高域側エネルギーＳＨとを用いて、信号の平均レベルｌａ情報を、

【００５２】

【数８】

報ｆｂと平均レベル情報ｌａの内のどちらか一つの情報からでも上述したように有声音か否かの判断が可能であるが、両方の情報を用いれば判断の信頼度は高くなる。すなわち、

$f_b < f_{thb}$　かつ　$l_a < l_{tha}$

のとき有声音でないという信頼度の高い判断ができる。そして、出力端子３８から該判断情報を導出する。

【００５５】ここで、この第２の実施例での上記エネル

ギー分布情報f_bと平均レベル情報l_aを別々に、上述した第１の実施例での時間軸上の標準偏差又は実効値の偏在情報ある比率（偏在情報）p_fと組み合わせて有声音か否かの判断を行うこともできる。すなわち、

$p_f < p_{thf}$ かつ $f_b < f_{thb}$ 又は$p_f < p_{thf}$ かつ $l_a < l_{tha}$

のとき有声音でないという信頼度の高い判断を行うことができる。

【００５６】以上により、この第２の実施例は、有声音のエネルギー分布が周波数軸上の低域側に集中し、雑音又は無声音のエネルギー分布が周波数軸上の高域側に集中する傾向を用いて有声音か否かを判別することができる。

【００５７】次に図６は、本発明の第３の実施例としての有声音判別方法を説明するための有声音判別装置の概略構成を示す図である。

【００５８】この図６において、入力端子４１には、少なくとも低域成分（２００Hz以下）が除去され、方形窓により１ブロックＮサンプル（例えばＮ＝２５６）で窓かけ処理されて時間軸方向に移動され、さらに１ブロックが細分割されたサブブロック毎の信号が供給される。このサブブロック毎の信号から上記統計的性質検出部１４が統計的性質を検出する。そして上記第１の実施例で説明したような偏在検出部１７又は１９が上記統計的性質から統計的性質の時間軸上での偏りを検出する。この偏在検出部１７又は１９からの偏在情報は、判断部３９に供給される。また、入力端子４２には、少なくとも低域成分（２００Hz以下）が除去され、ハミング窓により１ブロックＮサンプル（例えばＮ＝２５６）で窓かけ処理されて時間軸方向に移動され、さらに直交変換により周波数軸上に変換されたデータが供給される。この周波数軸上に変換されたデータは、上記エネルギー検出部３４に供給される。このエネルギー検出部３４により検出された高域側エネルギー検出値と低域側エネルギー検出値は、エネルギー分布算出部３５に供給される。このエネルギー分布計算部３５により求められたエネルギー分布情報は、判断部３９に供給される。さらに、上記高域側エネルギー検出値と低域側エネルギー検出値は、信号レベル算出部３６に供給され、１サンプル当たりの信号のレベルが計算される。この信号レベル計算部３６によって計算された信号レベル情報は、上記判断部３９に供給される。上記判断部３９には、上記偏在情報、エネルギー分布情報及び信号レベル情報が供給される。これらの情報により判断部３９は、入力音声信号が有声音であるか否かを判断する。そして、出力端子４３から該判断情報を導出する。

【００５９】以下に、この第３の実施例の動作を説明する。この第３の実施例は、上記偏在検出部１７、１９からの各サブフレーム毎の信号の偏向情報p_f、上記分布算出部３５からのエネルギー分布情報f_b及び上記信号

レベル算出部３６からの平均レベル情報l_aを用いて上記判断部３９で有声音か否かの判断を行うものである。例えば、

$p_f < p_{thf}$ かつ $f_b < f_{thb}$ かつ $l_a < l_{tha}$

のとき有声音でないという信頼度の高い判断を行う。

【００６０】以上により、この第３の実施例は、統計的性質の時間軸上での偏在情報、エネルギー分布情報及び平均レベル情報とに応じて有声音か否かを判断する。

【００６１】なお、本発明の上記実施例に係る有声音判別方法は、上記具体例にのみ限定されるものでないことはいうまでもない。例えば、各サブフレーム毎の信号の偏在情報p_fを用いて有声音を判別する場合には、その時間変化を追い例えば５フレーム連続して

$p_f < p_{thf}$　　（$p_{thf} = 1.1$）

のときに限りフラットとみなしフラグP_{fs}を１とする。一方、５フレームの内１フレームでも、

$p_f \geqq p_{thf}$

となったら、上記フラグP_{fs}を０とする。そして、

$f_b < f_{bt}$ かつ $P_{fs} = 1$ かつ $l_a < l_{tha}$

のときに有声音でないという信頼度の非常に高い判断を行うことができる。

【００６２】そして、本発明に係る有声音判別方法により、有声音でない、すなわち、背景雑音又は子音と判断されたときには、入力音声信号の１ブロックを全て強制的に無声音とすることにより、ＭＢＥ等のボコーダの合成側での異音の発生を防ぐことができる。

【００６３】次に、本発明に係る有声音判別方法の第４の実施例について、図７及び図８を参照しながら説明する。上述した第１の実施例においては、信号の上記サブブロック毎の標準偏差や実効値（ｒｍｓ値）のデータの分布を調べるために、標準偏差や実効値の各データの相加平均と相乗平均との比率を求めているが、上記相乗平均をとるためには、上記１フレーム内のサブブロックの個数（例えば３２個）のデータの乗算と３２乗根の演算とが必要とされる。この場合、先に３２個のデータを乗算するとオーバーフロー（桁あふれ）が生ずるため、先に各データのそれぞれ３２乗根をとった後に乗算を行うような工夫が必要とされる。このとき、３２個の各データ毎に３２回の３２乗根演算が必要となり、多くの演算量が要求されることになる。

【００６４】そこで、この第４の実施例においては、上記３２個の各サブブロック毎の実効値（ｒｍｓ値）のフレーム内での標準偏差σ_{rms}と平均値$\underline{rms}$とを求め、これらの値に応じて（例えばこれらの値の比率に応じて）実効値ｒｍｓの分布を検出している。すなわち、上記各サブブロック毎の実効値ｒｍｓ、このｒｍｓのフレーム内の標準偏差σ_{rms}及び平均値$\underline{rms}$は、

【００６５】

【数９】

15　　16

$$rms(i)=\sqrt{\frac{1}{B_L}\sum_{j=0}^{B_L-1}x^2\ (i\cdot B_L+j)}\qquad 0\leqq i<B_N\ (=32)\quad\cdots\ (14)$$

$$\underline{rms}\ =\ \frac{1}{B_N}\sum_{i=0}^{B_N-1}rms\ (i)\qquad (B_N=32)\quad\cdots\ (15)$$

$$\sigma_{rms}=\sqrt{\frac{1}{B_N}\sum_{i=0}^{B_N-1}(\ rms(i)\ -\underline{rms}\ )^2}\qquad\cdots\ (16)$$

【００６６】と表せる。これらの式中で、ｉは上記サブブロックのインデックス（例えばｉ＝０～３１）、B_Lはサブブロック内のサンプル数（サブブロック長、例えばB_L＝８）、B_Nは１フレーム内のサブブロックの個数（例えばB_N＝３２）をそれぞれ示し、１フレーム内のサンプル数Ｎを例えば２５６としている。

【００６７】上記（16）式の標準偏差σ_{rms}は、信号レベルが大きくなるとそれだけで大きくなってしまうので、上記（15）式の平均値ｒｍｓで割り込んで正規化（ノーマライズ）する。この正規化（ノーマライズ）した標準偏差をσ_mとするとき、

$$\sigma_m=\sigma_{rms}/rms\qquad\cdots\ (17)$$

となる。このσ_mは、有声部では大きな値となり、無声部又は背景雑音部分では小さな値となる。このσ_mが閾値σ_{th}より大きいときは有声とみなし、閾値σ_{th}より小さいときは無声又は背景雑音の可能性ありとして、他の条件（信号レベルやスペクトルの傾き）のチェックを行う。なお、上記閾値σ_{th}の具体的な値としては、σ_{th}＝０．４が挙げられる。

【００６８】以上のような時間軸上のエネルギー分布の分析処理は、図８のＡに示すような音声の母音部と図８のＢに示すようなノイズ又は音声の子音部とで、上記サブフレーム毎の短時間実効値（ｒｍｓ値）の分布に違いが見られることに着目したものである。すなわち、図８のＡの母音部での上記短時間ｒｍｓ値の分布（曲線ｂ参照）には大きな偏りがあるのに対して、図８のＢのノイズ又は子音部での短時間ｒｍｓ値の分布（曲線ｂ）はほぼフラットである。なお、図８のＡ、Ｂの各曲線ａは信号波形（サンプル値）を示している。このような短時間ｒｍｓ値の分布を調べるために、本実施例では、短時間ｒｍｓ値のフレーム内の標準偏差σ_{rms}と平均値ｒｍｓとの比率、すなわち上記正規化（ノーマライズ）された標準偏差をσ_mを用いているわけである。

【００６９】この時間軸上のエネルギー分布の分析処理のための構成については、図７の入力端子５１からの入力データを、実効値算出部６１に送って上記サブブロック毎の実効値ｒｍｓ(ｉ)を求め、平均値及び標準偏差算出部６２に送って上記平均値ｒｍｓ及び標準偏差σ_{rms}を求めた後、正規化標準偏差算出部６３に送って上記正規化した標準偏差σ_mを求めている。この正規化標準偏差σ_mは、ノイズ又は無声区間判別部６４に送ってい

【００７０】次に、スペクトルの傾きのチェックについて説明する。通常、有声音部分では、周波数軸上で低域にエネルギーが集中する。これに対して無声部又は背景雑音部では高域側にエネルギーが集中しやすい。そこで、高域側と低域側のエネルギーの比をとって、その値を雑音部か否かの評価尺度の１つとして使用する。すなわち、図７の入力端子５１からの１ブロック（１フレーム）内のｘ(ｎ)（０≦ｎ＜Ｎ、Ｎ＝２５６）に対して、窓かけ処理部５２にて適当な窓（例えばハミング窓）をかけ、ＦＦＴ（高速フーリエ変換）部５３でＦＦＴ処理を行って得た結果を、

Ｒｅ(ｊ)　　（０≦ｊ＜Ｎ／２）

Ｉｍ(ｊ)　　（０≦ｊ＜Ｎ／２）

とする。ただし、Ｒｅ(ｊ)はＦＦＴ係数の実部、Ｉｍ(ｊ)は同虚部である。また、Ｎ／２は規格化周波数のπに相当し、実周波数の４ｋＨｚ（ｘ(ｎ)は８ｋＨｚサンプリングのデータなので）に当たる。

【００７１】上記ＦＦＴ処理結果は、振幅算出部５４に送って振幅a_m(ｊ)を求めている。この振幅算出部５４は、上記第２の実施例のエネルギー検出部３４と同様な処理を行う部分であり、上記（９）式の演算が行われる。次に、この演算結果である振幅a_m(ｊ)がS_L、S_H、f_b算出部５５に送られ、この算出部５５において、上記エネルギー検出部３４内の低域側、高域側の各エネルギー検出部３４ａ、３４ｂでの演算、すなわち上記（10）式による低域側エネルギーS_Lの演算、及び上記（11）式による高域側エネルギーS_Hの演算が行われ、さらにこれらの比率であるエネルギーバランスを示

すパラメータ f_b （$=S_L/S_H$、上記（12）式参照）を求めている。この値が小さいときは高域側にエネルギーが片寄っていてノイズ又は子音である可能性が高い。このパラメータ f_b を上記ノイズ又は無声区間判別部６４に送っている。

【００７２】次に、上記第２の実施例の信号レベル算出部３６に相当する信号パワー算出部５６において、上記（13）式に示す信号の平均レベルあるいはパワー l_a を算出している。この信号レベルあるいは信号パワー l_a も上記ノイズ又は無声区間判別部６４に送っている。

【００７３】ノイズ又は無声区間判別部６４においては、上記各算出された値 σ_m、f_b、l_a に基づいてノイズ又は無声区間を判別する。この判別ための処理をＦ（・）と定義するとき、Ｆ（σ_m、f_b、l_a）の関数の具体例として次のようなものが挙げられる。

【００７４】先ず、第１の具体例として、

$f_b < f_{bth}$ かつ $\sigma_m < \sigma_{mth}$ かつ $l_a < l_{ath}$

ただし、f_{bth}、σ_{mth}、l_{ath} はいずれも閾値

の条件とすることが考えられ、この条件が満足されるとき、ノイズと判断し、全バンドＵＶ（無声音）とする。ここで、各閾値の具体的な値としては、$f_{bth}=15$、$\sigma_{mth}=0.4$、$l_{ath}=550$が挙げられる。

【００７５】次に、第２の例として、上記正規化標準偏差 σ_m の信頼度を向上するために、もう少し長時間の σ_m を観測することも考えられる。具体的には、Ｍフレーム連続して $\sigma_m < \sigma_{mth}$ のときに限り、時間軸上のエネルギー分布がフラットであると見なし、σ_m 状態フラグ σ_{state} をセット（$\sigma_{state}=1$）する。１フレームでも $\sigma_m \leqq \sigma_{mth}$ が出現したときには、上記 σ_m 状態フラグ σ_{state} をリセット（$\sigma_{state}=0$）する。そして、上記関数Ｆ（・）としては、

$f_b < f_{bth}$ かつ $\sigma_{state}=1$ かつ $l_a < l_{ath}$

のときにノイズあるいは無声と判断し、Ｖ／ＵＶフラグをオールＵＶとする。

【００７６】上記第２の例のように正規化標準偏差 σ_m の信頼度を高めた状態においては、信号レベル（信号パワー）l_a のチェックを不要としてもよい。この場合の関数Ｆ（・）としては、

$f_b < f_{bth}$ かつ $\sigma_{state}=1$

のときに、無声又はノイズと判断すればよい。

【００７７】以上説明したような第４の実施例によれば、ＤＳＰへのインプリメントが可能な程度の少ない演算量で、正確にノイズ（背景雑音）区間や無声区間を検出することが可能となり、背景雑音と判定された部分（フレーム）は強制的に全バンドをＵＶとすることで、背景雑音をエンコード／デコードすることによるうなり音のような異音の発生を抑えることが可能になる。

【００７８】以下、本発明に係る有声音判別方法が適用可能な音声信号の合成分析符号化装置（いわゆるボコーダ）の一種のＭＢＥ（Multiband Excitation: マルチバンド励起）ボコーダの具体例について、図面を参照しながら説明する。このＭＢＥボコーダは、D. W. Griffin and J. S. Lim, "Multiband Excitation Vocoder," IEEE Trans. Acoustics, Speech, and Signal Processing, vol. 36, No. 8, pp. 1223-1235, Aug. 1988に開示されているものであり、従来のＰＡＲＣＯＲ（PARtialauto-CORrelation: 偏自己相関）ボコーダ等では、音声のモデル化の際に有声音区間と無声音区間とをブロックあるいはフレーム毎に切り換えていたのに対し、ＭＢＥボコーダでは、同時刻（同じブロックあるいはフレーム内）の周波数軸領域に有声音（Voiced）区間と無声音（Unvoiced）区間とが存在するという仮定でモデル化している。

【００７９】図９は、上記ＭＢＥボコーダの実施例の全体の概略構成を示すブロック図である。この図９において、入力端子１０１には音声信号が供給されるようになっており、この入力音声信号は、ＨＰＦ（ハイパスフィルタ）等のフィルタ１０２に送られて、いわゆるＤＣ（直流）オフセット分の除去や帯域制限（例えば２００～３４００Hzに制限）のための少なくとも低域成分（２００Hz以下）の除去が行われる。このフィルタ１０２を介して得られた信号は、ピッチ抽出部１０３及び窓かけ処理部１０４にそれぞれ送られる。ピッチ抽出部１０３では、入力音声信号データが所定サンプル数Ｎ（例えばＮ＝２５６）単位でブロック分割され（あるいは方形窓による切り出しが行われ）、このブロック内の音声信号についてのピッチ抽出が行われる。このような切り出しブロック（２５６サンプル）を、例えば図１０のＡに示すようにＬサンプル（例えばＬ＝１６０）のフレーム間隔で時間軸方向に移動させており、各ブロック間のオーバラップはＮ－Ｌサンプル（例えば９６サンプル）となっている。また、窓かけ処理部１０４では、１ブロックＮサンプルに対して所定の窓関数、例えばハミング窓をかけ、この窓かけブロックを１フレームＬサンプルの間隔で時間軸方向に順次移動させている。

【００８０】このような窓かけ処理を数式で表すと、

$$x_w(k,q) = x(q)\, w(kL-q) \quad \cdots (18)$$

となる。この（18）式において、ｋはブロック番号を、ｑはデータの時間インデックス（サンプル番号）を表し、処理前の入力信号のｑ番目のデータｘ(q) に対して第ｋブロックの窓（ウィンドウ）関数ｗ(kL-q)により窓かけ処理されることによりデータ $x_w(k,q)$ が得られることを示している。ピッチ抽出部１０３内での図１０のＡに示すような方形窓の場合の窓関数 $w_r(r)$ は、

$$w_r(r) = 1 \quad (0 \leqq r < N)$$
$$= 0 \quad (r < 0,\ N \leqq r) \quad \cdots (19)$$

19

また、窓かけ処理部１０４での図１０のＢに示すような

$$w_h(r) = 0.54 - 0.46\cos(2\pi r/(N-1))$$

$$= 0$$

である。このような窓関数$w_r(r)$あるいは$w_h(r)$を用いるときの上記（18）式の窓関数w(r)（＝w(kL-q)）の否零区間は、

$0 \leqq kL - q < N$

これを変形して、

$kL - N < q \leqq kL$

従って、例えば上記方形窓の場合に窓関数$w_r(kL-q)=1$となるのは、図１１に示すように、$kL-N<q\leqq kL$のときとなる。また、上記（18）～（20）式は、長さＮ（＝２５６）サンプルの窓が、Ｌ（＝１６０）サンプルずつ前進してゆくことを示している。以下、上記（19）式、（20）式の各窓関数で切り出された各Ｎ点（$0 \leqq r < N$）の否零サンプル列を、それぞれ$x_{wr}(k,r)$、$x_{wh}(k,r)$と表すことにする。

【００８１】窓かけ処理部１０４では、図１２に示すように、上記（20）式のハミング窓がかけられた１ブロック２５６サンプルのサンプル列$x_{wh}(k,r)$に対して１７９２サンプル分の０データが付加されて（いわゆる０詰めされて）２０４８サンプルとされ、この２０４８サンプルの時間軸データ列に対して、直交変換部１０５により例えばＦＦＴ（高速フーリエ変換）等の直交変換処理が施される。

【００８２】ピッチ抽出部１０３では、上記$x_{wr}(k,r)$のサンプル列（１ブロックＮサンプル）に基づいてピッチ抽出が行われる。このピッチ抽出法には、時間波形の周期性や、スペクトルの周期的周波数構造や、自己相関関数を用いるもの等が知られているが、本実施例では、センタクリップ波形の自己相関法を採用している。このときのブロック内でのセンタクリップレベルについては、１ブロックにつき１つのクリップレベルを設定してもよいが、ブロックを細分割した各部（各サブブロック）の信号のピークレベル等を検出し、これらの各サブブロックのピークレベル等の差が大きいときに、ブロック内でクリップレベルを段階的にあるいは連続的に変化

$$S(j) = H(j) \cdot |E(j)|$$

と表現するようなモデルを想定している。ここで、Ｊは$\pi\omega_S = f_S/2$に対応し、サンプリング周波数$f_S = 2\pi\omega_S$が例えば８ｋＨｚのときには４ｋＨｚに対応する。上記（21）式中において、周波数軸上のスペクトルデータS(j)が図１３のＡに示すような波形のとき、H(j)は、図１３のＢに示すような元のスペクトルデータS(j)のスペクトル包絡線（エンベロープ）を示し、E(j)は、図１３のＣに示すような等レベルで周期的な励起信号（エキサイテイション）のスペクトルを示している。すなわち、ＦＦＴスペクトルS(j)は、スペクトルエンベロープH(j)と励起信号のパワースペクトル|E

20

ハミング窓の場合の窓関数$w_h(r)$は、

$0 \leqq r < N$　　・・・（20）

$r < 0,\ N \leqq r$

させるようにしている。このセンタクリップ波形の自己相関データのピーク位置に基づいてピーク周期を決めている。このとき、現在フレームに属する自己相関データ（自己相関は１ブロックＮサンプルのデータを対象として求められる）から複数のピークを求めておき、これらの複数のピークの内の最大ピークが所定の閾値以上のときには該最大ピーク位置をピッチ周期とし、それ以外のときには、現在フレーム以外のフレーム、例えば前後のフレームで求められたピッチに対して所定の関係を満たすピッチ範囲内、例えば前フレームのピッチを中心として±２０％の範囲内にあるピークを求め、このピーク位置に基づいて現在フレームのピッチを決定するようにしている。このピッチ抽出部１０３ではオープンループによる比較的ラフなピッチのサーチが行われ、抽出されたピッチデータは高精度（ファイン）ピッチサーチ部１０６に送られて、クローズドループによる高精度のピッチサーチ（ピッチのファインサーチ）が行われる。

【００８３】高精度（ファイン）ピッチサーチ部１０６には、ピッチ抽出部１０３で抽出された整数（インテジャー）値の粗（ラフ）ピッチデータと、直交変換部１０５により例えばＦＦＴされた周波数軸上のデータとが供給されている。この高精度ピッチサーチ部１０６では、上記粗ピッチデータ値を中心に、0.２～0.５きざみで±数サンプルずつ振って、最適な小数点付き（フローティング）のファインピッチデータの値へ追い込む。このときのファインサーチの手法として、いわゆる合成による分析（Analysis by Synthesis)法を用い、合成されたパワースペクトルが原音のパワースペクトルに最も近くなるようにピッチを選んでいる。

【００８４】このピッチのファインサーチについて説明する。先ず、上記ＭＢＥボコーダにおいては、上記ＦＦＴ等により直交変換された周波数軸上のスペクトルデータとしてのS(j)を

$0 < j < J$　　・・・（21）

(j)|との積としてモデル化される。

【００８５】上記励起信号のパワースペクトル|E(j)|は、上記ピッチに応じて決定される周波数軸上の波形の周期性（ピッチ構造）を考慮して、１つの帯域（バンド）の波形に相当するスペクトル波形を周波数軸上の各バンド毎に繰り返すように配列することにより形成される。この１バンド分の波形は、例えば上記図１２に示すような２５６サンプルのハミング窓関数に１７９２サンプル分の０データを付加（０詰め）した波形を時間軸信号と見なしてＦＦＴし、得られた周波数軸上のある帯域幅を持つインパルス波形を上記ピッチに応じて切り出す

ことにより形成することができる。

【００８６】次に、上記ピッチに応じて分割された各バンド毎に、上記H(j)を代表させるような（各バンド毎のエラーを最小化するような）値（一種の振幅）$|A_m|$を求める。ここで、例えば第mバンド（第m高調波の帯域）の下限、上限の点をそれぞれa_m、b_mとするとき、この第mバンドのエラーε_mは、

【００８７】

【数１０】

$$\varepsilon_m = \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} \quad \cdots (22)$$

【００８８】で表せる。このエラーε_mを最小化するような$|A_m|$は、

【００８９】

【数１１】

$$\frac{\partial \varepsilon_m}{\partial |A_m|} = 2 \sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \} |E(j)|$$

$$\therefore |A_m| = 2 \sum_{j=a_m}^{b_m} |S(j)| |E(j)| \Big/ \sum_{j=a_m}^{b_m} |E(j)|^2 \quad \cdots (23)$$

【００９０】となり、この（23）式の$|A_m|$のとき、エラーε_mを最小化する。このような振幅$|A_m|$を各バンド毎に求め、得られた各振幅$|A_m|$を用いて上記（22）式で定義された各バンド毎のエラーε_mを求める。次に、このような各バンド毎のエラーε_mの全バンドの総和値$\Sigma\varepsilon_m$を求める。さらに、このような全バンドのエラー総和値$\Sigma\varepsilon_m$を、いくつかの微小に異なるピッチについて求め、エラー総和値$\Sigma\varepsilon_m$が最小となるようなピッチを求める。

【００９１】すなわち、上記ピッチ抽出部１０３で求められたラフピッチを中心として、例えば 0.25 きざみで上下に数種類ずつ用意する。これらの複数種類の微小に異なるピッチの各ピッチに対してそれぞれ上記エラー総和値$\Sigma\varepsilon_m$を求める。この場合、ピッチが定まるとバンド幅が決まり、上記（23）式より、周波数軸上データのパワースペクトル$|S(j)|$と励起信号スペクトル$|E(j)|$とを用いて上記（22）式のエラーε_mを求め、その全バンドの総和値$\Sigma\varepsilon_m$を求めることができる。このエラー総和値$\Sigma\varepsilon_m$を各ピッチ毎に求め、最小となるエラー総和値に対応するピッチを最適のピッチとして決定するわけである。以上のようにして高精度ピッチサーチ部１０６で最適のファイン（例えば 0.25 きざみ）ピッチが求められ、この最適ピッチに対応する振幅$|A_m|$が決定される。

【００９２】以上ピッチのファインサーチの説明においては、説明を簡略化するために、全バンドが有声音（Voiced）の場合を想定しているが、上述したようにＭＢＥボコーダにおいては、同時刻の周波数軸上に無声音（Unvoiced）領域が存在するというモデルを採用していることから、上記各バンド毎に有声音／無声音の判別を行うことが必要とされる。

【００９３】上記高精度ピッチサーチ部１０６からの最適ピッチ及び振幅$|A_m|$のデータは、有声音／無声音判別部１０７に送られ、上記各バンド毎に有声音／無声音の判別が行われる。この判別のために、ＮＳＲ（ノイズｔｏシグナル比）を利用する。すなわち、第mバンドのＮＳＲは、

【００９４】

【数１２】

$$NSR = \frac{\sum_{j=a_m}^{b_m} \{ |S(j)| - |A_m| |E(j)| \}^2}{\sum_{j=a_m}^{b_m} |S(j)|^2} \quad \cdots (24)$$

【００９５】と表せ、このＮＳＲ値が所定の閾値（例えば0.3）より大のとき（エラーが大きい）ときには、そのバンドでの$|A_m| |E(j)|$による$|S(j)|$の近似が良くない（上記励起信号$|E(j)|$が基底として不適当である）と判断でき、当該バンドをＵＶ（Unvoiced、無声音）と判別する。これ以外のときは、近似がある程度良好に行われていると判断でき、そのバンドをＶ（Voiced、有声音）と判別する。

【００９６】次に、振幅再評価部１０８には、直交変換部１０５からの周波数軸上データ、高精度ピッチサーチ部１０６からのファインピッチと評価された振幅$|A_m|$との各データ、及び上記有声音／無声音判別部１０７からのＶ／ＵＶ（有声音／無声音）判別データが供給されている。この振幅再評価部１０８では、有声音／無声

音判別部１０７において無声音（ＵＶ）と判別されたバンドに関して、再度振幅を求めている。このＵＶのバンドについての振幅｜A_m｜$_{UV}$は、

【００９７】
【数１３】

$$|A_m|_{UV}=\sqrt{\sum_{j=a_m}^{b_m}|S(j)|^2/(b_m-a_m+1)} \quad \cdots(25)$$

【００９８】にて求められる。

【００９９】この振幅再評価部１０８からのデータは、データ数変換（一種のサンプリングレート変換）部１０９に送られる。このデータ数変換部１０９は、上記ピッチに応じて周波数軸上での分割帯域数が異なり、データ数（特に振幅データの数）が異なることを考慮して、一定の個数にするためのものである。すなわち、例えば有効帯域を３４００Ｈｚまでとすると、この有効帯域が上記ピッチに応じて、８バンド～６３バンドに分割されることになり、これらの各バンド毎に得られる上記振幅｜A_m｜（ＵＶバンドの振幅｜A_m｜$_{UV}$も含む）データの個数m_{MX}＋１も８～６３と変化することになる。このためデータ数変換部１０９では、この可変個数m_{MX}＋１の振幅データを一定個数N_C（例えば４４個）のデータに変換している。

【０１００】ここで本実施例においては、周波数軸上の有効帯域１ブロック分の振幅データに対して、ブロック内の最後のデータからブロック内の最初のデータまでの値を補間するようなダミーデータを付加してデータ個数をN_F個に拡大した後、帯域制限型のK_{OS}倍（例えば８倍）のオーバーサンプリングを施すことによりK_{OS}倍の個数の振幅データを求め、このK_{OS}倍の個数（（m_{MX}＋１）×K_{OS}個）の振幅データを直線補間してさらに多くのN_M個（例えば２０４８個）に拡張し、このN_M個のデータを間引いて上記一定個数N_C（例えば４４個）のデータに変換する。

【０１０１】このデータ数変換部１０９からのデータ（上記一定個数N_Cの振幅データ）がベクトル量子化部１１０に送られて、所定個数のデータ毎にまとめられてベクトルとされ、ベクトル量子化が施される。ベクトル量子化部１１０からの量子化出力データは、出力端子１１１を介して取り出される。また、上記高精度のピッチサーチ部１０６からの高精度（ファイン）ピッチデータは、ピッチ符号化部１１５で符号化され、出力端子１１２を介して取り出される。さらに、上記有声音／無声音判別部１０７からの有声音／無声音（Ｖ／ＵＶ）判別データは、出力端子１１３を介して取り出される。これらの各出力端子１１１～１１３からのデータは、所定の伝送フォーマットの信号とされて伝送される。

【０１０２】なお、これらの各データは、上記Ｎサンプル（例えば２５６サンプル）のブロック内のデータに対して処理を施すことにより得られるものであるが、ブロックは時間軸上を上記Ｌサンプルのフレームを単位として前進することから、伝送するデータは上記フレーム単位で得られる。すなわち、上記フレーム周期でピッチデータ、Ｖ／ＵＶ判別データ、振幅データが更新されることになる。

【０１０３】次に、伝送されて得られた上記各データに基づき音声信号を合成するための合成側（デコード側）の概略構成について、図１４を参照しながら説明する。この図１４において、入力端子１２１には上記ベクトル量子化された振幅データが、入力端子１２２には上記符号化されたピッチデータが、また入力端子１２３には上記Ｖ／ＵＶ判別データがそれぞれ供給される。入力端子１２１からの量子化振幅データは、逆ベクトル量子化部１２４に送られて逆量子化され、データ数逆変換部１２５に送られて逆変換され、得られた振幅データが有声音合成部１２６及び無声音合成部１２７に送られる。入力端子１２２からの符号化ピッチデータは、ピッチ復号化部１２８で復号化され、データ数逆変換部１２５、有声音合成部１２６及び無声音合成部１２７に送られる。また入力端子１２３からのＶ／ＵＶ判別データは、有声音合成部１２６及び無声音合成部１２７に送られる。

【０１０４】有声音合成部１２６では例えば余弦(cosine)波合成により時間軸上の有声音波形を合成し、無声音合成部１２７では例えばホワイトノイズをバンドパスフィルタでフィルタリングして時間軸上の無声音波形を合成し、これらの各有声音合成波形と無声音合成波形とを加算部１２９で加算合成して、出力端子１３０より取り出すようにしている。この場合、上記振幅データ、ピッチデータ及びＶ／ＵＶ判別データは、上記分析時の１フレーム（Ｌサンプル、例えば１６０サンプル）毎に更新されて与えられるが、フレーム間の連続性を高める（円滑化する）ために、上記振幅データやピッチデータの各値を１フレーム中の例えば中心位置における各データ値とし、次のフレームの中心位置までの間（合成時の１フレーム）の各データ値を補間により求める。すなわち、合成時の１フレーム（例えば上記分析フレームの中心から次の分析フレームの中心まで）において、先端サンプル点での各データ値と終端（次の合成フレームの先端）サンプル点での各データ値とが与えられ、これらのサンプル点間の各データ値を補間により求めるようにしている。

【０１０５】以下、有声音合成部１２６における合成処

理を詳細に説明する。上記Ｖ（有声音）と判別された第ｍバンド（第ｍ高調波の帯域）における時間軸上の上記１合成フレーム（Ｌサンプル、例えば１６０サンプル）

$$V_m(n) = A_m(n)\cos(\theta_m(n)) \qquad 0 \leqq n < L \quad \cdots (26)$$

と表すことができる。全バンドの内のＶ（有声音）と判別された全てのバンドの有声音を加算（$\Sigma V_m(n)$）して最終的な有声音Ｖ(n) を合成する。

【０１０６】この（26）式中の$A_m(n)$は、上記合成フレームの先端から終端までの間で補間された第ｍ高調波の振幅である。最も簡単には、フレーム単位で更新され

$$A_m(n) = (L-n)A_{0m}/L + nA_{Lm}/L \quad \cdots (27)$$

の式により$A_m(n)$を計算すればよい。

【０１０７】次に、上記（26）式中の位相$\theta_m(n)$は、

$$\theta_m(0) = m\omega_{01}n + n^2 m(\omega_{L1}-\omega_{01})/2L + \phi_{0m} + \Delta\omega n \quad \cdots (28)$$

により求めることができる。この（28）式中で、ϕ_{0m}は上記合成フレームの先端（ｎ＝０）での第ｍ高調波の位相（フレーム初期位相）を示し、ω_{01}は合成フレーム先端（ｎ＝０）での基本角周波数、ω_{L1}は該合成フレームの終端（ｎ＝Ｌ：次の合成フレーム先端）での基本角周波数をそれぞれ示している。上記（28）式中の$\Delta\omega$は、ｎ＝Ｌにおける位相ϕ_{Lm}が$\theta_m(L)$に等しくなるような最小の$\Delta\omega$を設定する。

【０１０８】以下、任意の第ｍバンドにおいて、それぞれｎ＝０、ｎ＝Ｌのときの Ｖ／ＵＶ判別結果に応じた上記振幅$A_m(n)$、位相$\theta_m(n)$の求め方を説明する。第ｍバンドが、ｎ＝０、ｎ＝ＬのいずれもＶ（有声音）とされる場合に、振幅$A_m(n)$は、上述した（27）式により、伝送された振幅値A_{0m}、A_{Lm}を直線補間して振幅$A_m(n)$を算出すればよい。位相$\theta_m(n)$は、ｎ＝０で

$$\theta_m(0) = \phi_{Lm} - m(\omega_{01}+\omega_{L1})L/2 \quad \cdots (29)$$

とし、かつ$\Delta\omega$＝０とする。

【０１１１】上記ｎ＝０、ｎ＝ＬのいずれもＶ（有声音）とされる場合に、$\theta_m(L)$がϕ_{Lm}となるように$\Delta\omega$を設定する手法について説明する。上記（24）式で、ｎ＝Ｌと置くことにより、

$$\begin{aligned}\theta_m(L) &= m\omega_{01}L + L^2 m(\omega_{L1}-\omega_{01})/2L + \phi_{0m} + \Delta\omega L \\ &= m(\omega_{01}+\omega_{L1})L/2 + \phi_{0m} + \Delta\omega L \\ &= \phi_{Lm}\end{aligned}$$

となり、これを整理すると、$\Delta\omega$は、

$$\Delta\omega = (\mathrm{mod}2\pi((\phi_{Lm}-\phi_{0m}) - mL(\omega_{01}+\omega_{L1})/2)/L \quad \cdots (30)$$

となる。この（30）式で$\mathrm{mod}2\pi(x)$とは、ｘの主値を$-\pi$～$+\pi$の間の値で返す関数である。例えば、ｘ＝１.３πのとき$\mathrm{mod}2\pi(x)$＝－０.７π、ｘ＝２.３πのとき$\mathrm{mod}2\pi(x)$＝０.３π、ｘ＝－１.３πのとき$\mathrm{mod}2\pi(x)$＝０.７π、等である。

【０１１２】ここで、図１５のＡは、音声信号のスペクトルの一例を示しており、バンド番号（ハーモニクスナンバ）ｍが８、９、１０の各バンドがＵＶ（無声音）とされ、他のバンドはＶ（有声音）とされている。このＶ（有声音）のバンドの時間軸信号が上記有声音合成部１２６により合成され、ＵＶ（無声音）のバンドの時間軸

分の有声音を$V_m(n)$とするとき、この合成フレーム内の時間インデックス（サンプル番号）ｎを用いて、

る振幅データの第ｍ高調波の値を直線補間すればよい。すなわち、上記合成フレームの先端（ｎ＝０）での第ｍ高調波の振幅値をA_{0m}、該合成フレームの終端（ｎ＝Ｌ：次の合成フレームの先端）での第ｍ高調波の振幅値をA_{Lm}とするとき、

$\theta_m(0) = \phi_{0m}$からｎ＝Ｌで$\theta_m(L)$がϕ_{Lm}となるように$\Delta\omega$を設定する。

【０１０９】次に、ｎ＝０のときＶ（有声音）で、ｎ＝Ｌのときｕｖ（無声音）とされる場合に、振幅$A_m(n)$は、$A_m(0)$の伝送振幅値A_{0m}から$A_m(L)$で０となるように直線補間する。ｎ＝Ｌでの伝送振幅値A_{Lm}は無声音の振幅値であり、後述する無声音合成の際に用いられる。位相$\theta_m(n)$は、$\theta_m(0) = \phi_{0m}$とし、かつ$\Delta\omega$＝０とする。

【０１１０】さらに、ｎ＝０のときＵＶ（無声音）で、ｎ＝Ｌのときｖ（有声音）とされる場合には、振幅$A_m(n)$は、ｎ＝０での振幅$A_m(0)$を０とし、ｎ＝Ｌで伝送された振幅値A_{Lm}となるように直線補間する。位相$\theta_m(n)$については、ｎ＝０での位相$\theta_m(0)$として、フレーム終端での位相値ϕ_{Lm}を用いて、

$$\theta_m(0) = \phi_{Lm} - m(\omega_{01}+\omega_{L1})L/2 \quad \cdots (29)$$

を設定する手法について説明する。上記（24）式で、ｎ＝Ｌと置くことにより、

$$\begin{aligned}\theta_m(L) &= m\omega_{01}L + L^2 m(\omega_{L1}-\omega_{01})/2L + \phi_{0m} + \Delta\omega L \\ &= m(\omega_{01}+\omega_{L1})L/2 + \phi_{0m} + \Delta\omega L \\ &= \phi_{Lm}\end{aligned}$$

となり、これを整理すると、$\Delta\omega$は、

$$\Delta\omega = (\mathrm{mod}2\pi((\phi_{Lm}-\phi_{0m}) - mL(\omega_{01}+\omega_{L1})/2)/L \quad \cdots (30)$$

信号が無声音合成部１２７で合成されるわけである。

【０１１３】以下、無声音合成部１２７における無声音合成処理を説明する。ホワイトノイズ発生部１３１からの時間軸上のホワイトノイズ信号波形を、所定の長さ（例えば２５６サンプル）で適当な窓関数（例えばハミング窓）により窓かけをし、ＳＴＦＴ処理部１３２によりＳＴＦＴ（ショートタームフーリエ変換）処理を施すことにより、図１５のＢに示すようなホワイトノイズの周波数軸上のパワースペクトルを得る。このＳＴＦＴ処理部１３２からのパワースペクトルをバンド振幅処理部１３３に送り、図１５のＣに示すように、上記ＵＶ（無

声音）とされたバンド（例えばm＝８、９、１０）について上記振幅｜A_m｜$_{UV}$を乗算し、他のＶ（有声音）とされたバンドの振幅を０にする。このバンド振幅処理部１３３には上記振幅データ、ピッチデータ、Ｖ／ＵＶ判別データが供給されている。バンド振幅処理部１３３からの出力は、ＩＳＴＦＴ処理部１３４に送られ、位相は元のホワイトノイズの位相を用いて逆ＳＴＦＴ処理を施すことにより時間軸上の信号に変換する。ＩＳＴＦＴ処理部１３４からの出力は、オーバーラップ加算部１３５に送られ、時間軸上で適当な（元の連続的なノイズ波形を復元できるように）重み付けをしながらオーバーラップ及び加算を繰り返し、連続的な時間軸波形を合成する。オーバーラップ加算部１３５からの出力信号が上記加算部１２９に送られる。

【０１１４】このように、各合成部１２６、１２７において合成されて時間軸上に戻された有声音部及び無声音部の各信号は、加算部１２９により適当な固定の混合比で加算して、出力端子１３０より再生された音声信号を取り出す。

【０１１５】なお、上記図５の音声分析側（エンコード側）の構成や図１４の音声合成側（デコード側）の構成については、各部をハードウェア的に記載しているが、いわゆるＤＳＰ（ディジタル信号プロセッサ）等を用いてソフトウェアプログラムにより実現することも可能である。

【０１１６】また、本発明に係る有声音判別方法は、例えば、自動車電話の送信側で環境雑音（背景雑音等）を落としたいというようなとき、背景雑音を検出する手段としても用いられる。すなわち、雑音に乱された低品質の音声を処理し、雑音の影響を取り除き、聞きやすい音にするようないわゆるスピーチエンハンスメントでの雑音検出にも適用される。

【０１１７】

【発明の効果】本発明に係る有声音判別方法は、信号の１ブロックをさらに分割した複数のサブブロック毎に求めた信号の統計的な性質の時間軸上での偏りに応じて有声音を雑音又は無声音かと区別することにより、確実に判別できる。そして、ＭＢＥ等のボコーダに適用する場合には、音声のサブブロックに有声音入力がないとき、すなわち雑音又は無声音の入力があるとき、強制的に入力音声信号の全帯域を無声音として、間違ったピッチを検出することがないようにし、合成側での異音の発生を抑えることができる。

【０１１８】また、サブブロック毎の実効値（短時間ｒｍｓ値）の標準偏差及び平均値に基づいて短時間ｒｍｓ値の分布を調べることにより、少ない演算量で正確な有声音区間判別が行える。

【図面の簡単な説明】

【図１】本発明に係る有声音判別方法の第１の実施例を説明するための有声音判別装置の概略構成を示す機能ブロック図である。

【図２】信号の統計的性質を説明するための波形図である。

【図３】第１の実施例を説明するための有声音判別装置の要部の構成を示す機能ブロック図である。

【図４】第１の実施例を説明するための有声音判別装置の要部の構成を示す機能ブロック図である。

【図５】本発明に係る有声音判別方法の第２の実施例を説明するための有声音判別装置の概略構成を示す機能ブロック図である。

【図６】本発明に係る有声音判別方法の第３の実施例を説明するための有声音判別装置の要部の概略構成を示す機能ブロック図である。

【図７】本発明に係る有声音判別方法の第４の実施例を説明するための有声音判別装置の概略構成を示す機能ブロック図である。

【図８】信号の統計的性質としての短時間ｒｍｓ値の分布を説明するための波形図である。

【図９】本発明に係る有声音判別方法が適用可能な装置の具体例としての音声信号の合成分析符号化装置の分析側（エンコード側）の概略構成を示す機能ブロック図である。

【図１０】窓かけ処理を説明するための図である。

【図１１】窓かけ処理と窓関数との関係を説明するための図である。

【図１２】直交変換（ＦＦＴ）処理対象としての時間軸データを示す図である。

【図１３】周波数軸上のスペクトルデータ、スペクトル包絡線（エンベロープ）及び励起信号のパワースペクトルを示す図である。

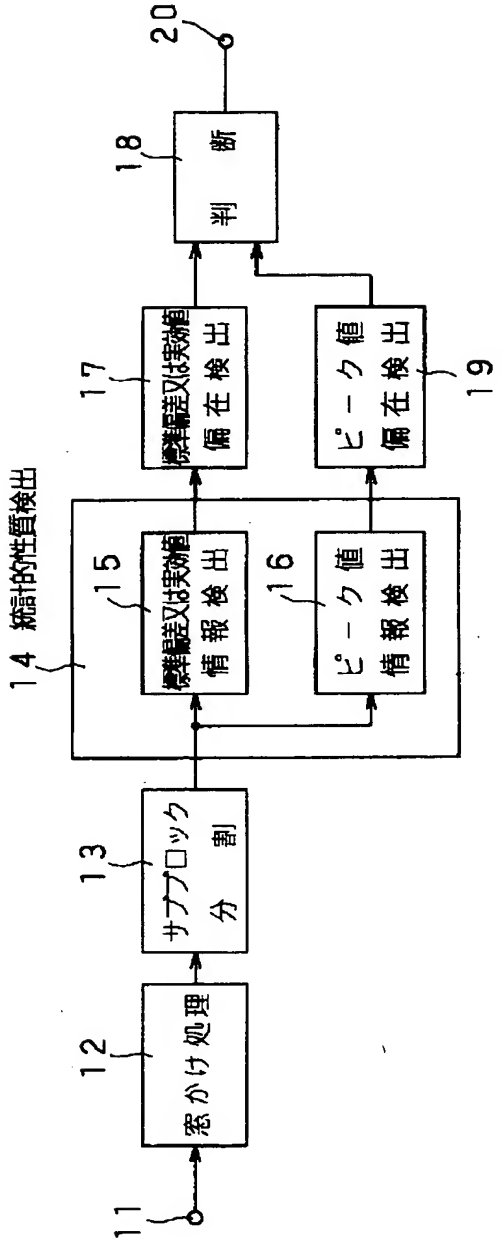
【図１４】本発明に係る有声音判別方法が適用可能な装置の具体例としての音声信号の合成分析符号化装置の合成側（デコード側）の概略構成を示す機能ブロック図である。

【図１５】音声信号を合成する際の無声音合成を説明するための図である。

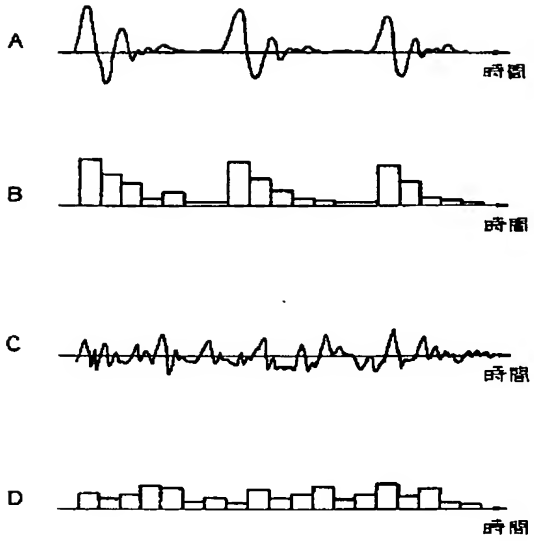
【符号の説明】

１２・・・・・窓かけ処理部
１３・・・・・サブブロック分割部
１４・・・・・統計的性質検出部
１５・・・・・標準偏差又は実効値情報検出部
１６・・・・・ピーク値情報検出部
１７・・・・・標準偏差又は実効値偏在検出部
１８・・・・・判断部
１９・・・・・ピーク値偏在検出部
６１・・・・・サブブロック毎の実効値算出部
６２・・・・・実効値の平均と標準偏差算出部
６３・・・・・正規化された標準偏差算出部

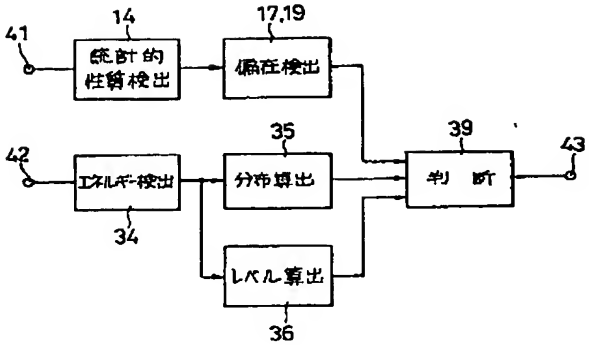
【図１】


11
12
窓かけ処理
13
サブブロック分割
14 統計的性質検出
15
標準偏差又は実効値情報検出
16
ピーク値情報検出
17
標準偏差又は実効値偏在検出
19
ピーク値偏在検出
18
判断
20

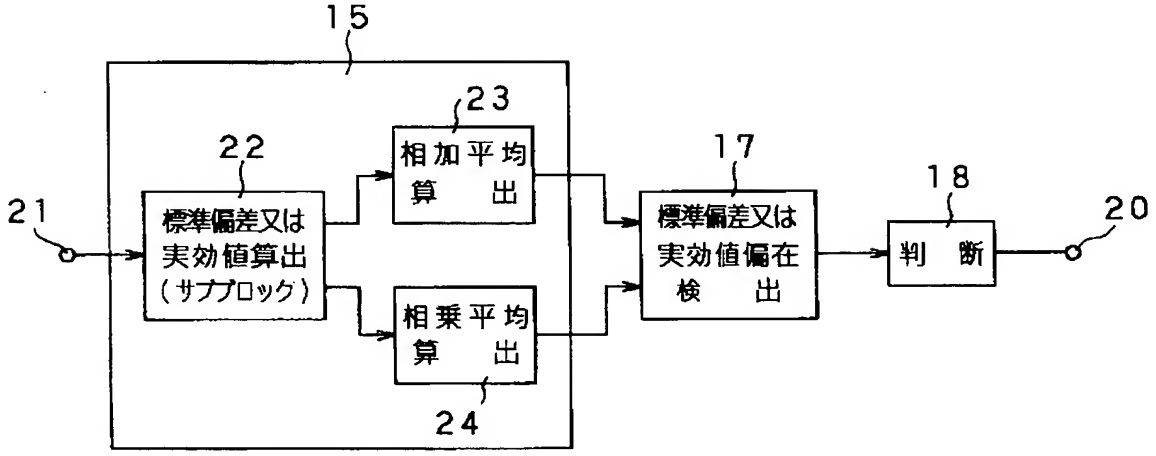
【図２】


A
時間
B
時間
C
時間
D
時間

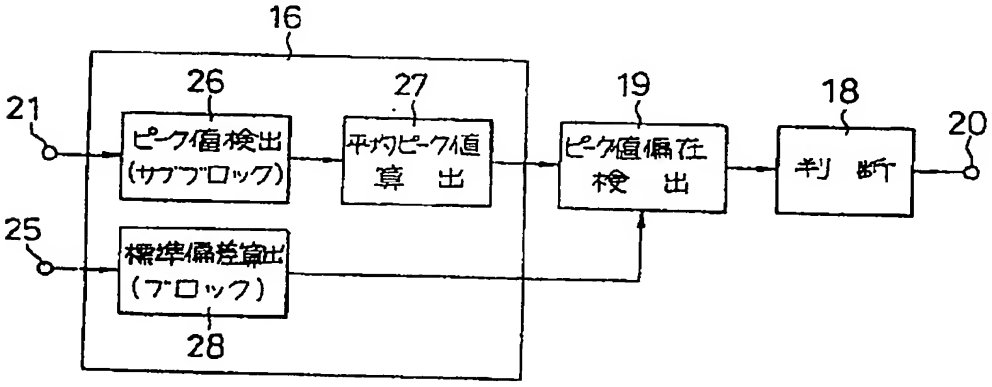
【図６】


41
14
統計的性質検出
17,19
偏在検出
42
エネルギー検出
34
35
分布算出
36
レベル算出
39
判　断
43

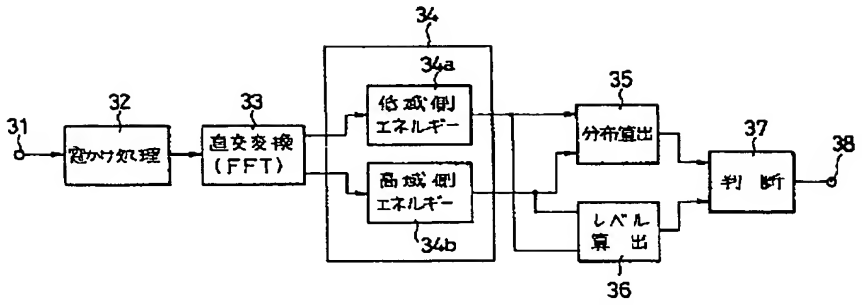
【図３】


15
21
22
標準偏差又は実効値算出（サブブロック）
23
相加平均算　出
24
相乗平均算　出
17
標準偏差又は実効値偏在検　出
18
判　断
20

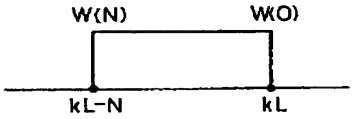
【図４】


16
21
26
ピーク値検出（サブブロック）
27
平均ピーク値算　出
25
標準偏差算出（ブロック）
28
19
ピーク値偏在検　出
18
判　断
20

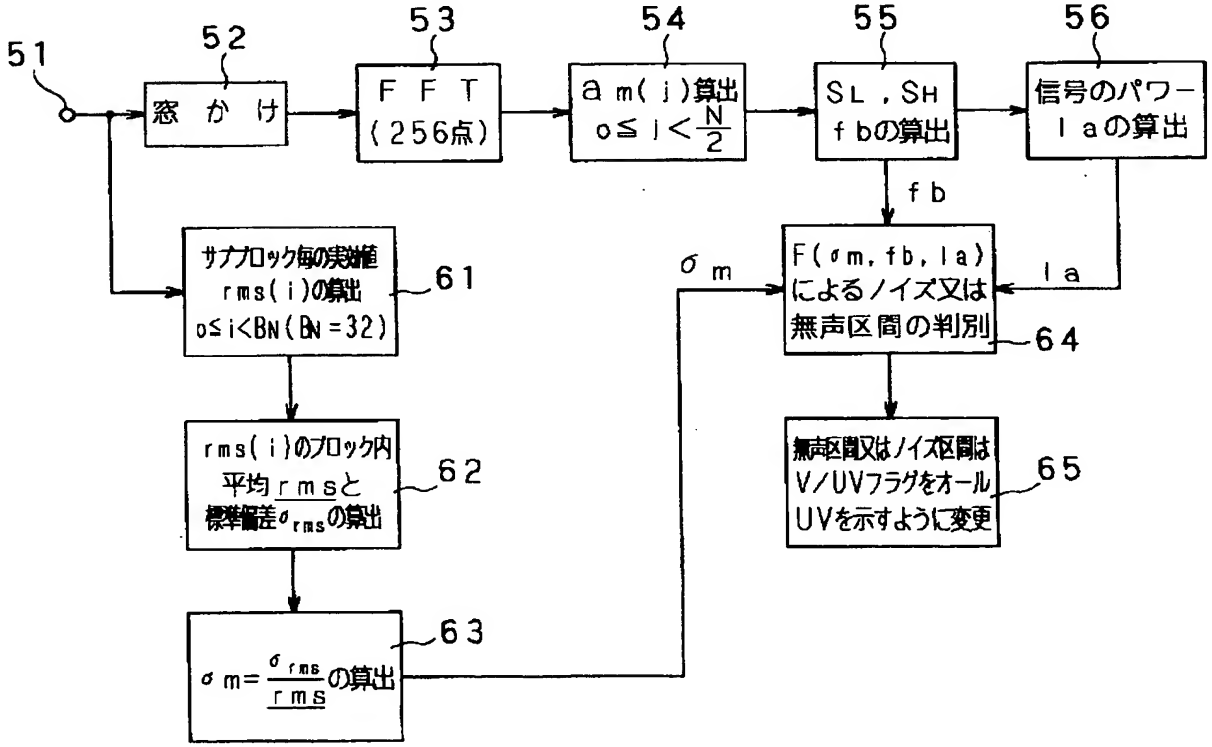
【図５】



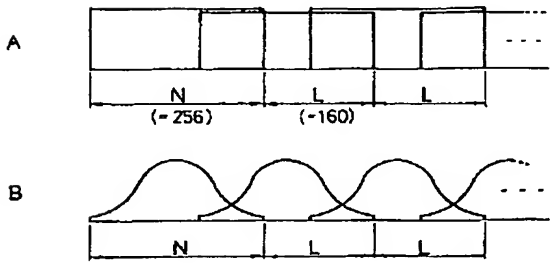
【図１１】



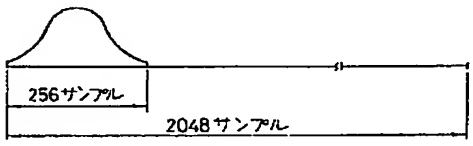
【図７】



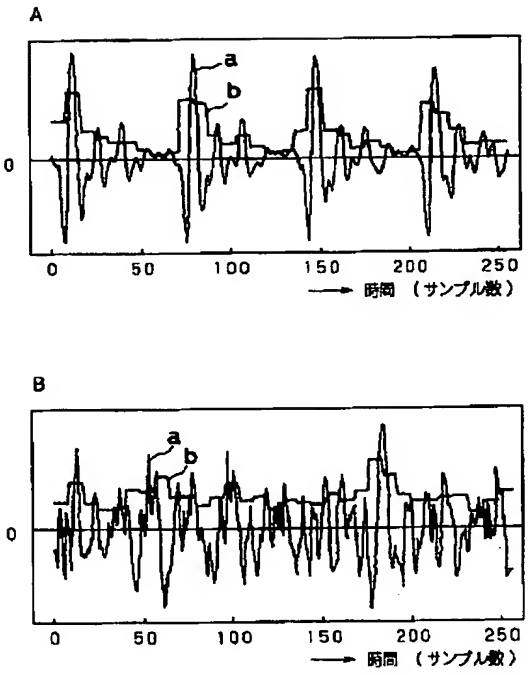
【図１０】



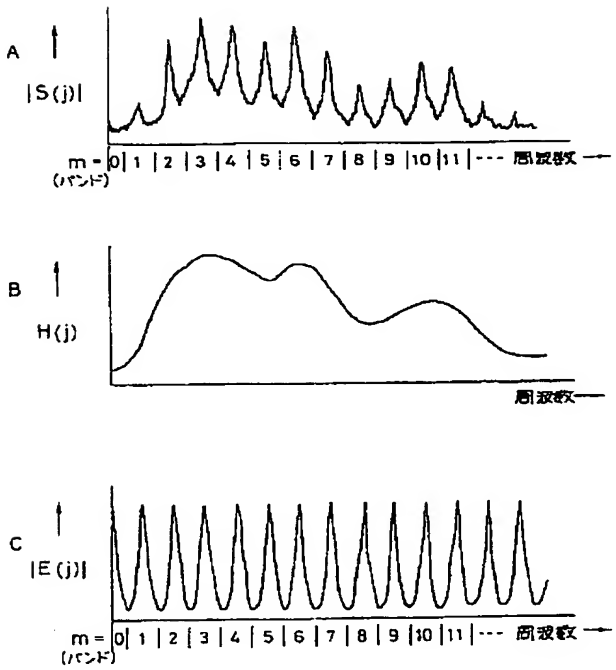
【図１２】



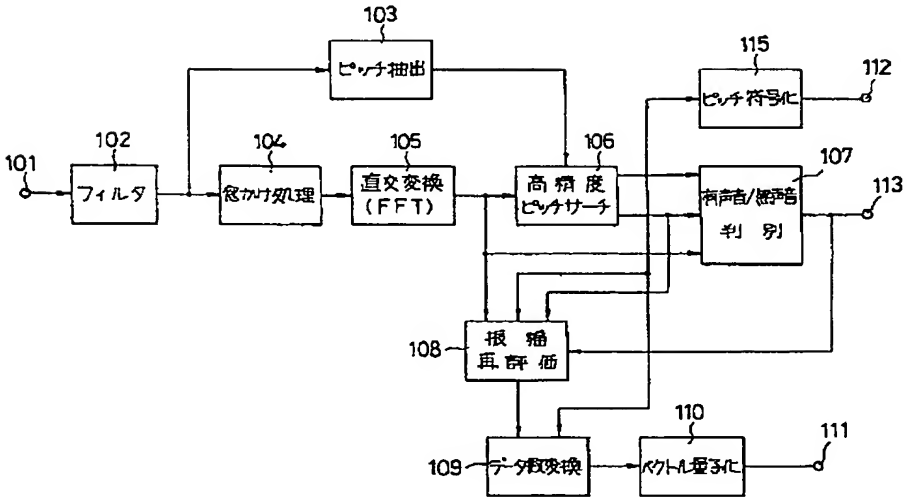
【図８】


A
a
b
0
0
50
100
150
200
250
時間（サンプル数）
B
a
b
0
0
50
100
150
200
250
時間（サンプル数）

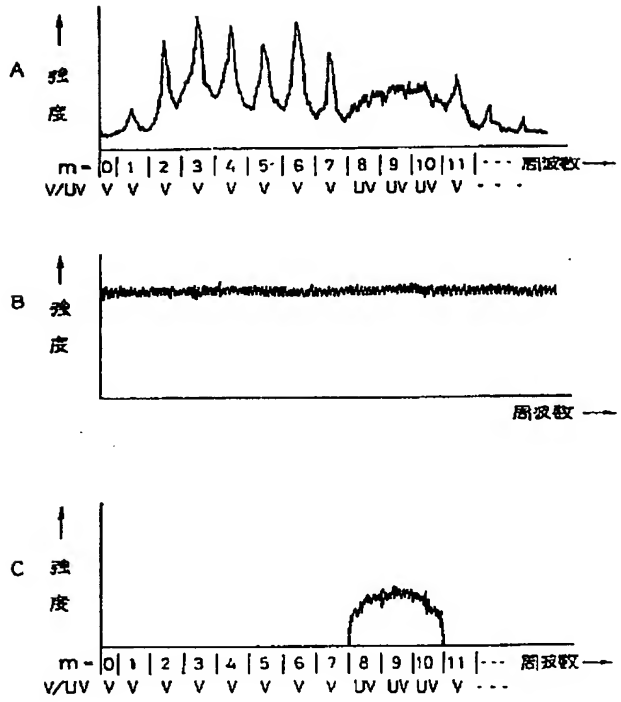
【図１５】


A
|S(j)|
m =
(バンド)
0 1 2 3 4 5 6 7 8 9 10 11 --- 周波数
B
H(j)
周波数
C
|E(j)|
m =
(バンド)
0 1 2 3 4 5 6 7 8 9 10 11 --- 周波数

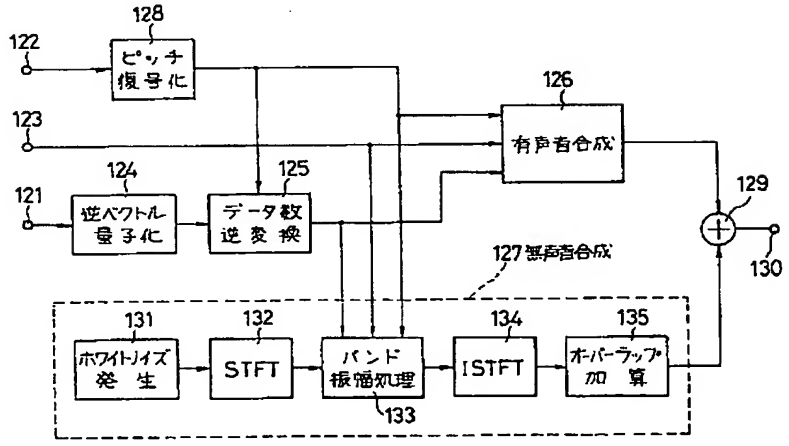
【図９】


103
ピッチ抽出
115
ピッチ符号化
112
102
104
105
106
101
フィルタ
窓かけ処理
直交変換
(FFT)
高精度
ピッチサーチ
107
有声音/無声音
判別
113
108
振幅
再評価
110
109
データ数変換
ベクトル量子化
111

【図１３】


A
強度
m=
0 1 2 3 4 5 6 7 8 9 10 11 ・・・ 周波数
V/UV V V V V V V V V UV UV UV V ・・・
B
強度
周波数
C
強度
m=
0 1 2 3 4 5 6 7 8 9 10 11 ・・・ 周波数
V/UV V V V V V V V V UV UV UV V ・・・

【図１４】


128
122
ピッチ復号化
126
123
有声音合成
124
125
121
逆ベクトル量子化
データ数逆変換
129
130
127無声音合成
131
132
134
135
ホワイトノイズ発生
STFT
バンド振幅処理
ISTFT
オーバーラップ加算
133